



TODD WARCZAK

Data Scientist

CONTACT

twarczak@gmail.com 

(253) 999-6478 

White River Junction, VT 

toddwarczak.netlify.app 

[LinkedIn](#) 

[Github](#) 

@TWarczak 

EDUCATION

Ph.D.

Molecular & Cellular Biology

Dartmouth College

September 2012 - August 2020

Hanover, NH

B.S.

Biology

University of Utah

September 2007 - May 2012

Salt Lake City, UT

SKILLS

R {tidyverse} {tidymodels}

SQL

Github

Markdown

Data Cleaning/Wrangling

Data Visualization

Probabilities/Statistics

Time-Series Forecasting

Machine Learning

WORK EXPERIENCE

Molecular Biologist

Dartmouth College

September 2012 - September 2020 / Hanover, NH

- Engineered genome-wide association study (GWAS) to identify genes controlling arsenic tolerance in plants. Determined variation in the expression of Arabidopsis gene NIP1;1 on 4th chromosome as the major genetic factor for tolerating arsenic. All data analyzed in R.
- Developed first cell-type specific expression map in R for plant genes involved in arsenic acquisition, efflux, sequestration, etc. for root epidermal, cortical, and endodermal cells.
- Built first RNA-seq pipeline for research lab to analyze gene expression of 25000+ plant genes using R. Transitioned multiple lab data analysis workflows from excel to R, including PCA, qPCR, Gene Ontology analysis, and statistical tests (t-tests, ANOVA, etc.).
- Presented research to local community members, government officials, and other stakeholders as a representative of the Dartmouth Toxic Metals Superfund Research Program.
- Mentored 2 undergraduate scientists.

PROJECTS

Forecasting Daily Sales with {modeltime}

- Forecasted 3 months of daily sales utilizing the {modeltime} package in R to help 'Superstore' company selling furniture, technology, and office supplies manage supply-chain/inventory decisions in Q1 (data: [Kaggle](#)).
- Exploratory data analysis and 11 models tested. 6 {tidymodels} workflows tuned for individual forecasts plus weighted ensemble (Support Vector Machine/Neural Network-AR/Random Forest/Prophet-XGboost) forecast. [Blog](#).

TidyTuesday

- Weekly twitter project focusing on cleaning, wrangling, summarizing, and arranging a new dataset in R to produce a single chart. Typically using {ggplot2} and {tidyverse} tools. Shared via #TidyTuesday. [Github](#).
- Mario Kart 64 World Records - Visualization of cumulative days individual records are held by players.
- Survivor - Visualization showing Myers-Briggs Type Indicator personalities of all show contestants and only the winners.